

Aanzet tot een nationale URI-Strategie voor Linked Data van de Nederlandse overheid

Dit document is een kopie van de versie van 24 september 2013 van het artikel op <http://www.pilod.nl/wiki/Boek/URI-strategie> in printbare vorm. Eventuele updates van latere datum zijn niet opgenomen.

Over dit document

Auteurs

- Hans Overbeek (KOOP) [1]
- Linda van den Brink (Geonovum) [2]

Vertaling

De inhoud van dit artikel is ook beschikbaar in een Engelse vertaling: [Towards a national URI-Strategy for Linked Data of the Dutch public sector](#) (pdf, EN).

Totstandkoming

Dit artikel is geschreven naar aanleiding van de Pilot Linked Open Data (PiLOD) van Geonovum in 2012 en 2013 [3]. Het verwoordt de inzichten die de 'Werkgroep URI-strategie' binnen de actielijn 'Techniek' heeft opgedaan. De werkgroep is ingesteld om te onderzoeken of een nationale URI-strategie voor Linked Data van de overheid zinvol is en zo ja, wat de ingrediënten voor zo'n strategie zouden moeten zijn. De leden van de werkgroep concluderen dat er bij tal van uitvoerende instanties behoefte bestaat aan richtlijnen voor het opstellen van URIs. Linked Data is een nieuw vakgebied, kent de nodige complexiteit en behelst domein-overstijgende vraagstukken. Technische, maar ook organisatorische. Dit document is nog geen URI-Strategie, maar geeft een inventarisatie van de issues en enkele aanbevelingen.

Belangrijke kennisbronnen die we in de werkgroep gebruikt hebben om te komen tot deze aanzet tot een nationale URI-strategie zijn:

- De Inspire richtlijn, die een nationale strategie voorschrijft voor URIs voor geo-informatie met de aanbeveling om deze geo-strategie te verbinden met een generieke nationale strategie [4]
- Designing URI sets for the UK Public Sector. Een aanbeveling van de Britse overheid die voorlopers zijn op het gebied van het publiceren van Linked Open Overheids-Data [5]
- 10 Rules for persistent URIs. Een veelomvattend rapport van de EU met vergelijkbare initiatieven en een waardevol overzicht van de laatste best-practices [6]

Tijdens de W3C workshop 'Open Data on the Web' [7] waren tal van internationale experts op het gebied van Linked Data aanwezig. Wij hebben daar een paper [8] gepresenteerd [9] met onze inzichten en vragen. Dit heeft waardevolle feedback opgeleverd, met name van Jeni Tennison [10] die heeft bijgedragen aan de Linked Data-richtlijnen van de Britse overheid. Hieruit blijkt al hoe belangrijk het is om Europese en andere internationale ontwikkelingen te volgen. Door deel te nemen in Europese en internationale programma's krijgen we toegang tot internationale kennis en expertise, blijven we op de hoogte van de laatste inzichten en toekomstige ontwikkelingen en kunnen we zelfs invloed uitoefenen op ontwikkeling van standaarden zodat deze zich ontwikkelen in lijn met de Nederlandse behoefte.

De werkgroep URI-strategie heeft geprobeerd om de kennis uit bovenstaande bronnen samen te vatten en toe te passen op de Nederlandse situatie. We kwamen in de werkgroep op sommige punten tot andere inzichten die deels ook bevestigd werden door de ervaringen die inmiddels in de UK zijn opgedaan.

Vervolg

Dit artikel zal als stabiele versie, als eindproduct van de PiLOD beschikbaar blijven. Hooguit zullen fouten worden verbeterd.

Wij hopen dat deze aanzet een vervolg krijgt in een daadwerkelijke URI-strategie die door beleidskaders wordt ondersteund. Tijdens de slotbijeenkomst van de PiLOD is dit artikel aangeboden aan Nico Westpalm van Hoorn, voorzitter van het Forum Standaardisatie [\[11\]](#) met het verzoek om bij de beleidsmakers aandacht te schenken aan de ontwikkelingen op het gebied van Linked Data en het belang van een overheidbreed gedragen URI-strategie.

Werkgroep

De auteurs willen alle deelnemers in de Werkgroep URI-strategie bedanken, zonder wie deze aanzet tot een nationale URI-strategie niet tot stand had kunnen komen:

- Thijs Brentjens (Geonovum)
- Wilko Quak (Geonovum)
- Paul Hermans (ProXML)
- Hayo Schreijer (KOOOP)
- Michel Grothe (Geonovum)
- Marcel van Mackelenbergh (Belastingdienst)
- Jan Jelle Boomgaardt (Digimelding)
- Rob van Dort (Mapplica)
- Arjen Santema (Kadaster)
- Bart van Leeuwen (Brandweer Amsterdam-Amstelland)
- Marco Brattinga (Kadaster)

en andere deelnemers aan de PiLOD.

Achtergrond

Linked Data

Over de toegevoegde waarde van Linked (Open) Data voor de overheid en de BV Nederland is al veel gezegd. De belangrijkste baten hiervan zijn efficiëntere bedrijfsvoering door de overheid (kostenreductie en sneller handelen); betrouwbaarheid en transparantie van de overheid (duidelijkheid en verantwoording), en economische waarde van data.

In de praktijk blijft de opbrengst echter vaak achter bij de beloften of verwachtingen. Veelgehoorde klachten zijn dat de data slecht vindbaar is en niet gemakkelijk in samenhang kan worden gebracht met andere data. Deze problemen worden bestreden door data van anderen te kopiëren, met hoge kosten voor verzamelen, converteren en synchroniseren, of door het bouwen van dure landelijke voorzieningen. Het resultaat is een veelheid aan kopieën en veel twijfel over authenticiteit van gegevens.

Een betere oplossing is om de authentieke data duurzaam beschikbaar te maken zodat iedereen deze kan gebruiken. Daarvoor is het nodig dat de data voorzien wordt van betrouwbare identificatie, zodat je er naar kunt verwijzen en ook de verwijzingen van iemand anders begrijpt. Deze directe verwijzingen naar authentieke gegevens zorgen voor meer samenhang en betere vindbaarheid en maken kopiëren en synchroniseren overbodig. Deze manier van data koppelen wordt Linked Data genoemd. [12] De Basisregistraties zijn al opgezet om als authentieke bron te dienen voor referentiegegevens. Wat ontbreekt is een goede strategie om deze bronnen ook voor de machine leesbaar te ontsluiten. Er is behoefte aan een soort klittenband waarmee datasets moeiteloos aan elkaar kunnen worden geplakt.

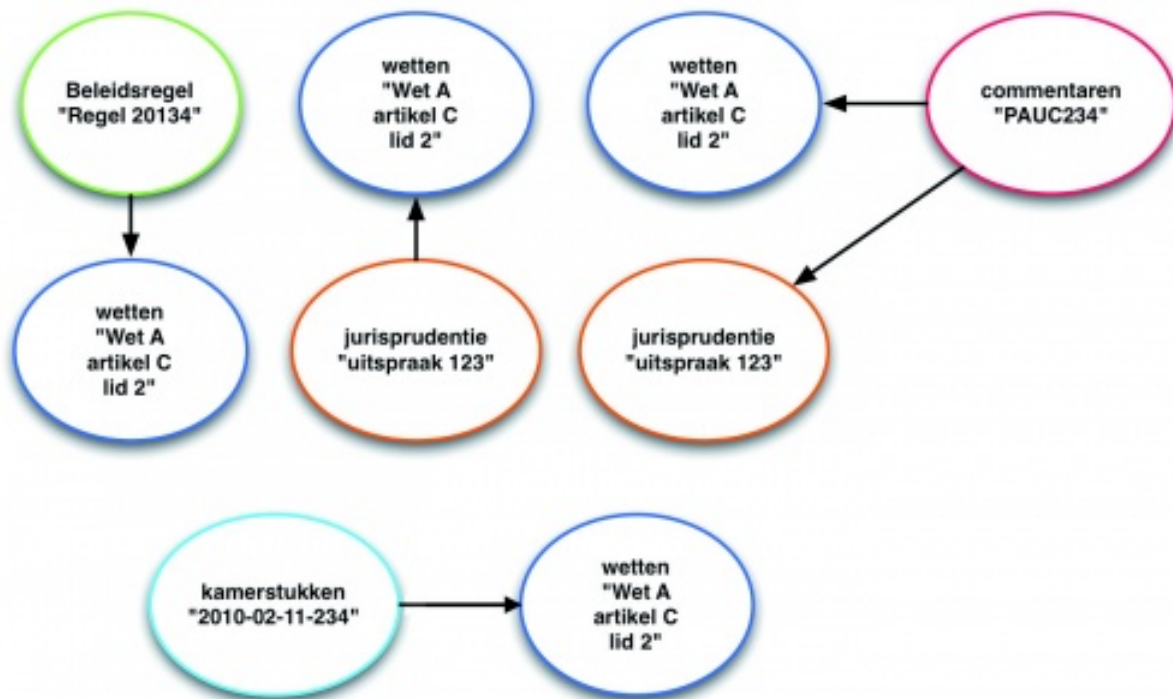
Het doel van de Pilot Linked Open Data (PiLOD) [13] van Geonovum was vooral om te onderzoeken wat er nodig is om van 'Open Data' bruikbare 'Linked Data' te maken. In de PiLOD is de behoefte geconstateerd aan een strategie voor identificatie van authentieke overheidsdata, te beginnen met de basisregistraties. Een aanzet voor zo'n strategie is gemaakt in de Werkgroep 'URI-strategie'. URI staat voor 'Uniform Resource Identifier', een standaard voor identificatie van objecten en concepten ('alles is een resource') van het W3C [14]. Centraal in deze strategie staan de basisregistraties en andere authentieke bronnen voor referentiegegevens, zoals bijvoorbeeld de collecties voor wet- en regelgeving op overheid.nl. Identificaties van authentieke referentiegegevens vormen namelijk de 'links' waar Linked Data mee gemaakt wordt.

Voorbeeld

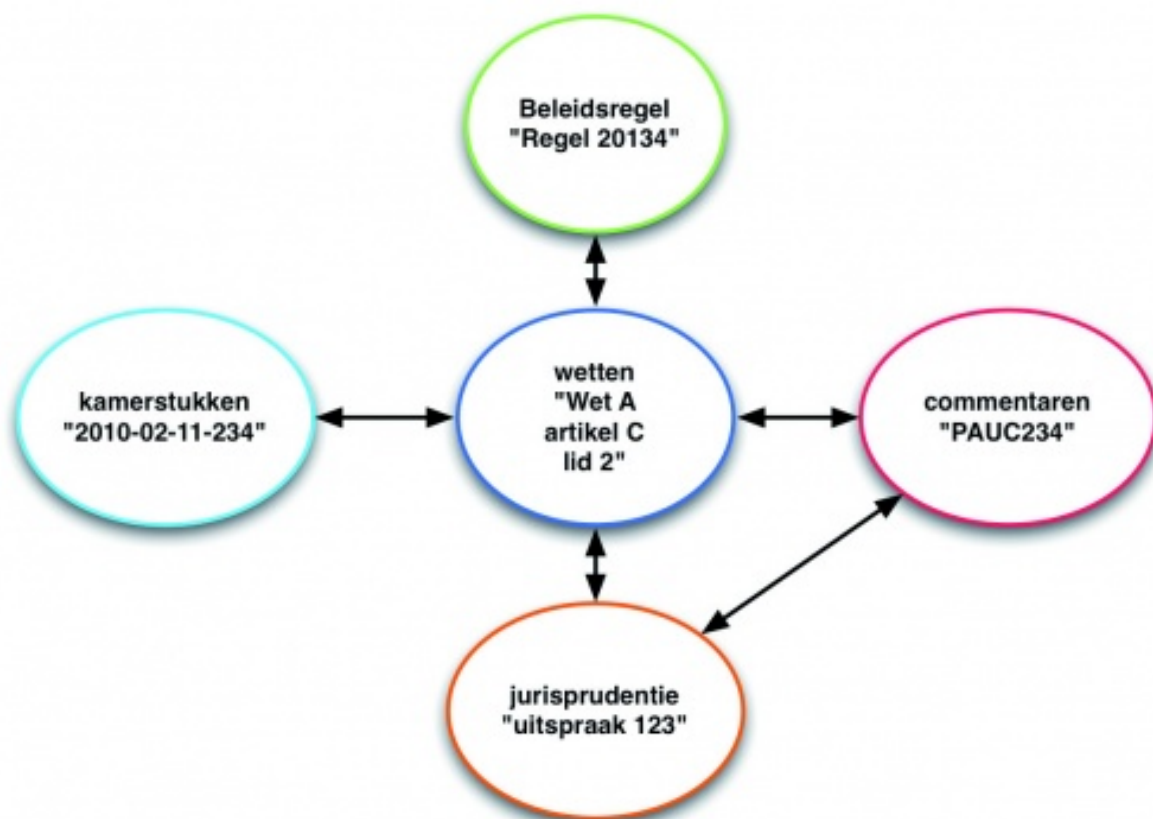
Laten we dit illustreren met een voorbeeld:

- Stel dat een uitvoeringsorganisatie een beleidsregel opstelt voor het uitvoeren van beleid. Deze beleidsregels hebben een wettelijke grondslag waarvoor de uitvoeringsorganisatie verwijst naar het betreffende lid in een artikel van de wet. De beleidsregel wordt gepubliceerd op hun website.
- Een rechter doet een uitspraak in een rechtszaak en baseert die uitspraak op datzelfde artikel. De uitspraak wordt gepubliceerd op rechtspraak.nl.
- Een rechtsgeleerde schrijft een commentaar op deze uitspraak en verwijst naar hetzelfde wetsartikel. Haar commentaar wordt gepubliceerd in een juridisch tijdschrift.

- Tenslotte wordt het wetsartikel besproken in de Tweede Kamer en de handelingen worden gepubliceerd op de parlementaire website.



Door nu de verwijzing naar het wetsartikel te standaardiseren kunnen de collecties van beleidsregels, uitspraken, commentaren en kamerstukken eenvoudig aan elkaar gelinkt worden. We kunnen nu bij elk van deze informatieobjecten eenvoudig een aantal gerelateerde documenten uit andere collecties aanreiken.



Het voorbeeld laat zien dat een uniforme identificatie van de referentie-objecten (de URIs) - in dit voorbeeld: wetsartikelen en uitspraken - essentieel is om de links tussen de informatieobjecten te kunnen leggen. Het zijn de haakjes en oogjes van het 'klittenband' waarmee de informatiecollecties aan elkaar gekoppeld worden.

Er is nog een andere vergelijking te maken, met de IBAN-nummers in de financiële wereld. Na jaren van problemen in het internationale en interbancaire geldverkeer, voeren financiële instellingen nu eindelijk een standaard door voor identificatie van bankrekeningen. Dit is een langdurige en kostbare operatie. Voor databanken dreigen vergelijkbare problemen: er worden op tal van plaatsen koppelingen gemaakt tussen data-collecties en iedere koppeling wordt opnieuw bedacht. We kunnen nu echter tijdig een strategie ontwikkelen om de objecten uit onze authentieke databanken (registers) een standaard 'bankrekeningnummer' te geven, zodat we problemen in een vroeg stadium kunnen oplossen, of zelfs voor zijn, en later kostbare herstructurering voorkomen.

Toegevoegde waarde van een URI-strategie

Een heldere strategie, in samenspraak met de stakeholders opgesteld, moet ervoor zorgen dat partijen die met Linked Data aan de slag willen verantwoorde keuzes kunnen maken die nodig zijn om Linked Data-oplossingen te maken. Idealiter wordt de URI-strategie, gericht op de technische implementatie van de identificatie van authentieke data, ingebed in een bredere Linked Data Strategie, waarin ook organisatorische aspecten worden meegenomen.

Een goede strategie heeft toegevoegde waarde voor meerdere al lopende trajecten.

- Stelselcatalogus 2.0 [15] stapt af van het idee van 1 overkoepelend model voor alle basisregistraties en kiest voor het in kaart brengen van de – onvermijdelijke – verschillen die er zijn tussen de modellen van de respectievelijke basisregistraties. Linked Data blijkt hier veel geschikter voor dan traditionele methoden voor datamodellering.
- OWMS, de metadatastandaard voor overheidsinformatie op het web ondersteunt zowel tekstwaarden (labels, bijvoorbeeld creator='Utrecht'), als het gebruik van identificers (pointers naar meer informatie, bijvoorbeeld creator=[http://standaarden.overheid.nl/owms/terms/Utrecht_\(gemeente\)](http://standaarden.overheid.nl/owms/terms/Utrecht_(gemeente))). Het gebruik van identificers levert veel nauwkeuriger verwijzingen op en biedt meer mogelijkheden dan labels.
- Data-collecties van de overheid [16] die als Linked Data worden gepubliceerd kunnen worden gelinkt met andere datasets en daardoor vaker en beter gebruikt dan datasets waar niet naar kan worden gelinkt. Die laatste moeten worden gekopieerd ze om te kunnen hergebruiken.
- In het Linked Data Overheid (LiDO) traject bij KOOP wordt wet- en regelgeving als bindmiddel voor Linked Data gebruikt. Hiermee wordt de samenhang en vindbaarheid van overheidsinformatie vergroot, wat contentintegratie eenvoudiger maakt.

De URI-strategie

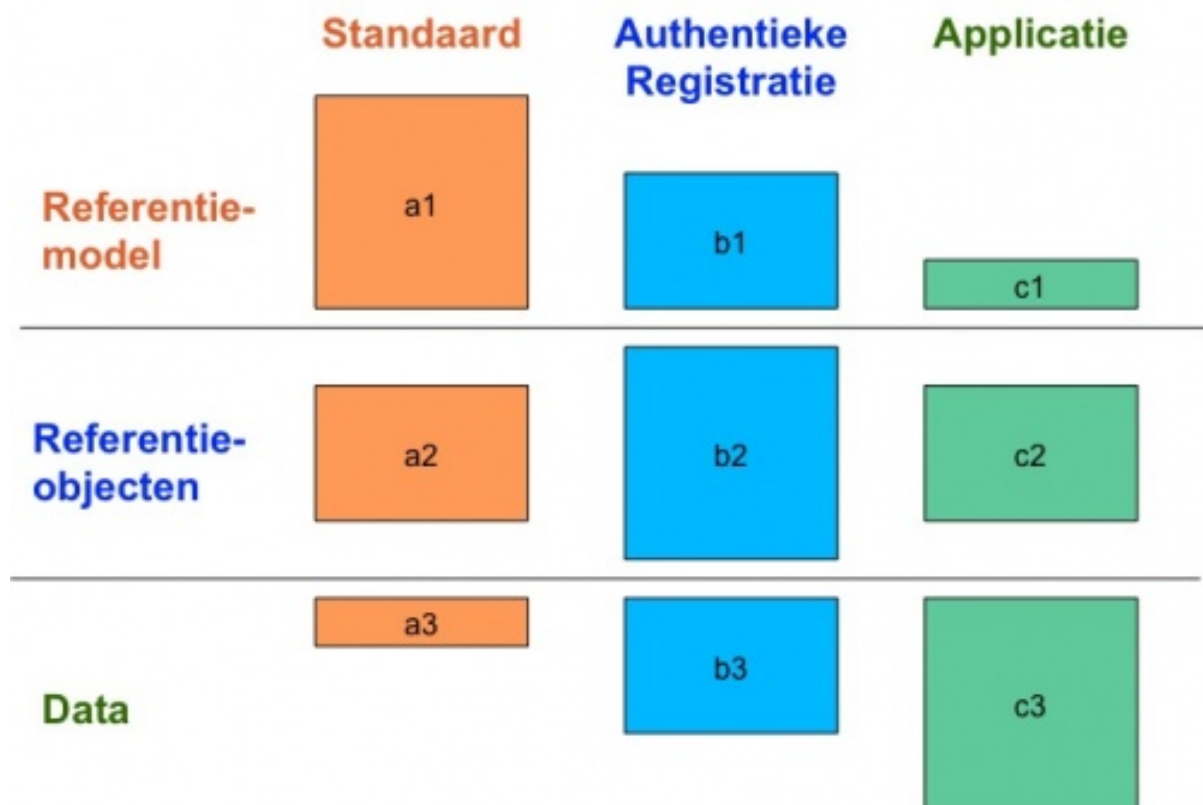
Scope

De URI-strategie is met name bedoeld voor data waarmee objecten of concepten worden gedefinieerd, waar andere toepassingen naar kunnen verwijzen. Data waar niet naar wordt gelinkt valt buiten de scope. Om dit toe te lichten onderscheiden we drie categorieën informatiebronnen:

1. Standaarden
2. Authentieke registraties
3. 'gewone' Applicaties

Elk met de nadruk op een van drie categorieën concepten:

1. Begrippen in een conceptueel Model
2. Referentieobjecten
3. 'gewone' Data

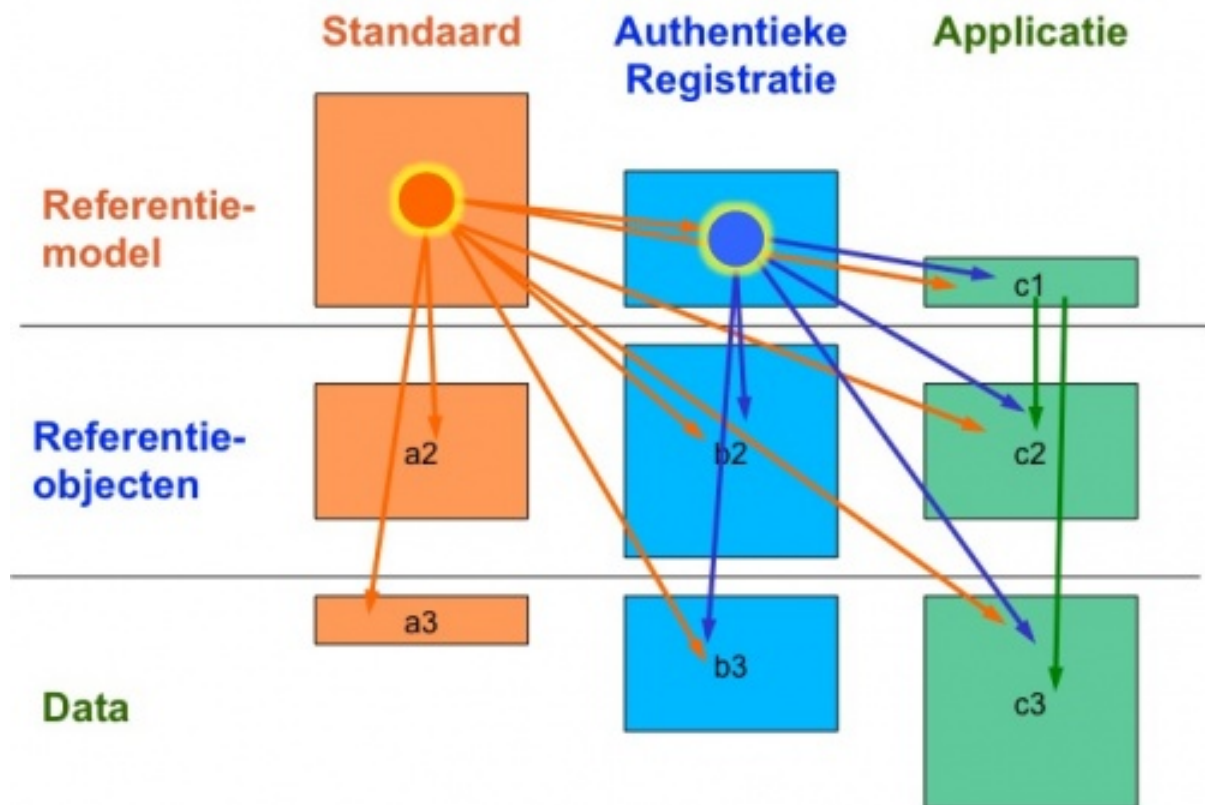


De grootte van een cel in het diagram geeft een indicatie van het gewicht van de categorie concepten in die categorie van informatiebronnen. De belangrijkste functie van een standaard is gewoonlijk om een conceptueel model te definiëren. (a1) Authentieke registraties worden doorgaans opgezet om een administratie van Referentieobjecten bij te houden (b2) en een 'gewone' Applicatie heeft gewoonlijk slechts de functie om Data te verzamelen voor een specifiek doel (c3).

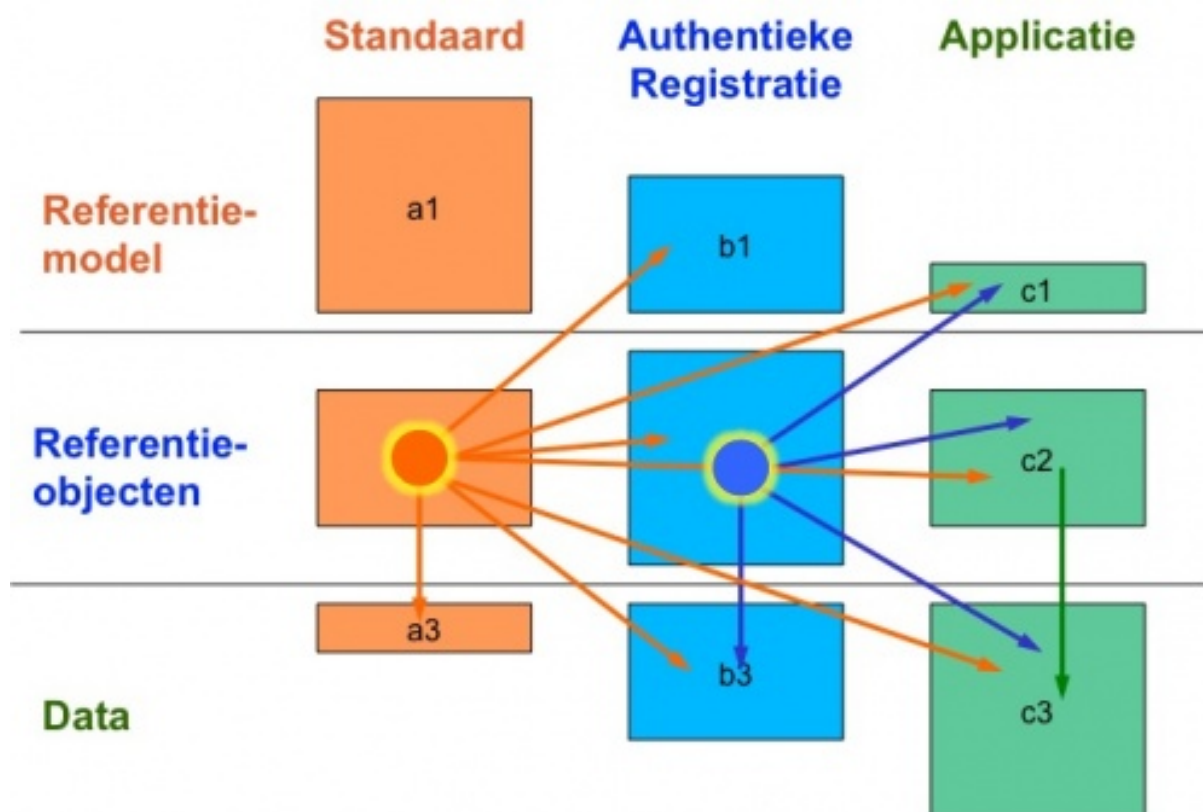
Uiteraard kunnen een Authentieke registratie en een Applicatie een eigen Model hebben (b1 en c1), sommige Standaarden verstrekken een lijst met Referentiewaarden (a2) en Applicaties kunnen lokale

Referentiegegevens hebben (c2). Ook kunnen Standaarden en Registers wat 'gewone' data nodig hebben (a3 en b3), bijvoorbeeld om wijzigingen en herkomst van hun referentiegegevens vast te leggen.

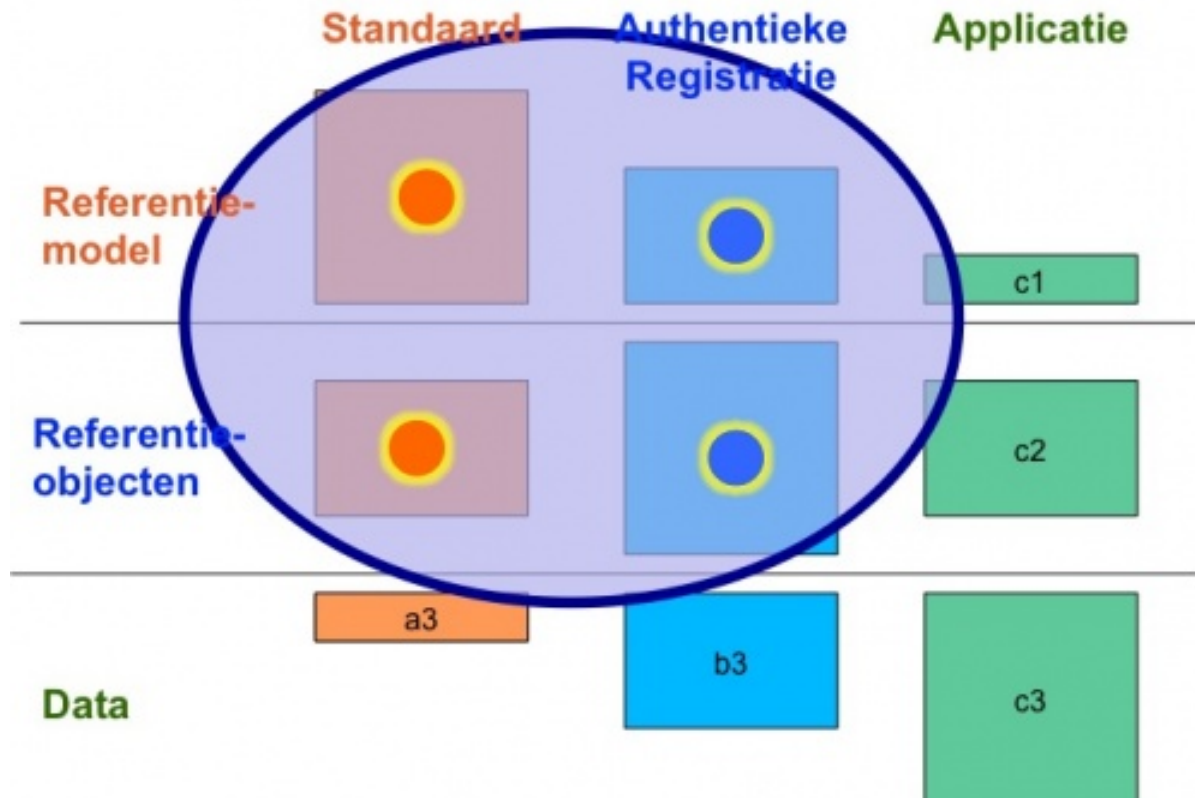
De URI-strategie ondersteunt het hergebruik van Concepten en Referentieobjecten door andere data-collecties. De interessante categorieën zijn dan ook de termen in de Modellen en de Referentiegegevens.



Termen, zoals klassen en eigenschappen, die zijn gedefinieerd in de Modellen van een Standaard of een Authentieke registratie, worden gebruikt om Referentieobjecten en Data te classificeren.



Referentieobjecten, gedefinieerd door Standaarden (denk aan waardenlijsten), maar vooral die beheerd worden in Authentieke registraties, worden gebruikt in 'gewone' Applicaties.



De URI-strategie is bedoeld voor Modellen en Referentieobjecten van zowel Standaarden als Authentieke registraties. Dus niet in eerste instantie voor de 'gewone' Data (rij 3) of concepten in 'gewone' Applicaties (kolom c) aangezien daar nou eenmaal minder naar wordt gelinkt.

Inzichten

Gedurende de PiLOD heeft de Werkgroep URI-strategie een aantal inzichten opgedaan bij de analyse van de voorgestelde varianten.

'No register, no identifier'

In Linked Data theorie wordt nogal eens beweerd dat het nodig is om voor elk concept of object een URI te definiëren, waardoor het lijkt alsof je pas kunt beginnen als je voor elk concept of object een nieuwe Linked Data URI hebt verzonnen en 'gemunt' (to mint a URI). Maar waarom zou je alles opnieuw definiëren? De mensheid definieert al eeuwen lang authentieke identificatie voor standaard begrippen en referentieobjecten. Denk aan encyclopedieën, taxonomieën en registraties van inwoners of van onroerende goederen. We noemen hier een voorziening voor het authentiek definiëren en identificeren van concepten of referentieobjecten een register. Onder register verstaan we hier dus zowel een specificatie van begrippen/concepten in een standaard, als een authentieke registratie van referentieobjecten.

Doel van deze niet geringe inspanning om registers op te zetten is, om vanuit verschillende administraties op een eenduidige manier met een afgesproken term (een identifier) naar nauwkeurige en meer uitgebreide definities van abstracte begrippen en objecten te kunnen verwijzen zodat iedereen weet wat bedoeld wordt. Informatie uit verschillende administraties kan vervolgens gekoppeld worden door – handmatig – gelijke termen (gewoonlijk namen of nummers) bij elkaar te zetten.

Nu we dat willen gaan automatiseren is er niets op tegen om deze bestaande registers te blijven gebruiken. Maar wat nou, als we naar begrippen of objecten willen verwijzen waar nog geen register voor bestaat? De enige manier om voor ontbrekende begrippen en objecten een URI te munten is toch door deze vast te leggen in een nieuw register. Als we merken dat er voor bepaalde begrippen of objecten geen register bestaat, terwijl we hier wel URIs voor willen hebben, dan is de enige manier om een register aan te leggen. Kortom: je kunt alleen URIs munten voor begrippen of objecten die vastliggen in een register. Dit belangrijke inzicht vatten we samen in het adagium: 'No register, No identifier'.

Geen verplicht gemeenschappelijk internetdomein

Het W3C beveelt aan om http-URIs te gebruiken om Linked Data mee te identificeren. Een http-URI is een URL en begint dus met 'http://', gevolgd door een domein en eventueel een lokaal pad. De Britse URI-strategie gaat er vanuit dat alle Linked Data URIs van de overheid worden ondergebracht op één hoofddomein: 'data.gov.uk'. Om dat schaalbaar te houden stellen zij voor om dat domein onder te verdelen in sectoren. Sectoren die genoemd worden zijn bijvoorbeeld: location, education, transport en health. De bijbehorende domeinen zijn dan: 'location.data.gov.uk', 'education.data.gov.uk', 'transport.data.gov.uk' en 'health.data.gov.uk'. Dat levert echter twee problemen op:

1. Voor elke sector moet een partij gevonden worden die eigenaar en beheerder van dat sub-domein wil zijn.
2. Niet alle informatie valt duidelijk in één domein. Vallen treinstations onder 'location' of onder 'transport'?

De sector-eigenaren zijn dus moeilijk te bepalen en zullen onderling verschillen in aangeboden diensten en gevolgde procedures. Maar er is bovendien een eigenaar van het nationale hoofddomein nodig. Daarmee zouden we zowel een single-point-of-failure als een organisatorische afhankelijkheid tussen de registerhouders en de houder van het centrale domein en de sectordomeinen creëren.

Vandaar dat we deze UK-richtlijn niet willen overnemen in de Nederlandse URI-strategie. In de UK onderkent men deze problemen inmiddels zelf ook. Bovendien zien zij ook problemen met het gebruik van het hoofddomein 'data.gov.uk'. Een van de adviezen die we in Londen kregen tijdens ODW13 was dan ook om elk domein dat beoogd onderdeel van duurzame URIs is, bij wet vast te leggen.

Uitgangspunten

De Werkgroep URI-strategie heeft een aantal uitgangspunten geformuleerd die gehanteerd zouden moeten worden bij het opstellen van de strategie:

1. **Sluit aan bij internationale best-practices.** Alleen ga je sneller, maar samen kom je verder. Door aan te sluiten op internationale ontwikkelingen profiteer je van oplossingen die wereldwijd bedacht worden. Ook is Europese regelgeving van steeds groter belang voor de Nederlandse overheid.
2. **Sluit aan bij bestaande ontwikkelingen.** De strategie raakt vele partijen en systemen en kan niet in een keer als iets nieuws worden ingevoerd. Kijk daarom goed naar wat er op het gebied van standaardisatie en authentieke registraties toch al gebeurt en maak daar maximaal gebruik van.
3. **Houd rekening met afwijkende strategieën.** Ook als er systemen worden gemaakt die om wat voor reden dan ook niet de nationale strategie volgen, moet hiermee gelinkt kunnen worden.
4. **Houd het zo simpel mogelijk, maar niet simpeler.** Bij een te complexe benadering zal de strategie niet of onvoldoende worden toegepast, bij een te eenvoudige benadering zal de strategie niet voldoende opleveren.

Denk bij standaardisatie goed na over:

Persistentie

Persistentie betekent dat oplossingen ook stand houden als de organisatie eromheen wijzigt. Ook al moeten we accepteren dat we nog niet alles weten en dat voortschrijdend inzicht tot andere keuzes kan leiden. Persistentie betekent niet voor de eeuwigheid, maar een onderneming of instantie moet er wel bedrijfskritische systemen op durven te ontwikkelen.

Schaalbaarheid

Schaalbaarheid is van belang om beheerkosten te kunnen blijven overzien, ook als toepassingen groeien. Het is onvoorspelbaar hoeveel applicaties er de komende jaren zullen ontstaan. Elk onderdeel van de strategie zal dan ook schaalbaar moeten worden opgezet.

Begrijpelijkheid.

Begrijpelijkheid is noodzakelijk om te zorgen dat afspraken makkelijk worden opgepakt en overgenomen.

Vertrouwen.

Vertrouwen is nodig om organisaties te bewegen om zelf strategisch te kiezen voor het gebruik en publicatie van Linked Data.

Machine-leesbaarheid.

Machine-leesbaarheid zorgt ervoor dat er met Linked Data ook werkende oplossingen kunnen worden gebouwd.

Menselijke leesbaarheid.

Menselijke leesbaarheid is ook belangrijk om te zorgen dat men oplossingen vertrouwt en begrijpt. Maar als de machine de data niet goed gebruiken kan, dan werkt het überhaupt niet.

...en bij voorkeur in die volgorde.

URI-patroon

In navolging van de drie eerder genoemde bronnen gaan wij er vanuit dat http-URLs de voor de hand liggende keuze vormen. In alle drie de strategieën wordt uitgegaan van nadere afspraken over het te gebruiken patroon om de http-URL op te bouwen. Het patroon voor http-URLs dat in deze bronnen wordt aanbevolen - en wat wij daarom overnemen - is:

```
http://{domain}/{type}/{concept}/{reference}
```

We behandelen de vier onderdelen hierna één voor één.

{domain}

Het {domain} deel bevat het internet domein en eventueel een pad binnen dat domein:

```
{domain} = {internet domain}/{path}.
```

Het {domain} dient twee doelen. Ten eerste is het een belangrijk instrument om unieke identificaties te verkrijgen: twee objecten, die beheerd worden in twee verschillende databases, kunnen toevallig dezelfde identificatie krijgen (bijvoorbeeld een kadastraal perceel met id 010101 en een rechtspersoon met id 010101). Als nu zowel het Kadaster als het Nieuw HandelsRegister (NHR) besluit om deze objecten als linked data te publiceren, worden er toch twee unieke URLs gevormd: de een begint bijvoorbeeld met http://brk.nl/ en de ander met http://nhr.nl/. Ten tweede zorgt een goedgekozen domein voor herkenbaarheid en vertrouwen. Kadastrale percelen met een URI als http://data.brk.nl/perceel/010101 hebben een betrouwbaarder uitstraling dan bijvoorbeeld http://data.vindhethier.eu/perceel/010101.

Het {path} kan worden gebruikt als binnen een register verschillende verzamelingen objecten leven, waarbij dubbele id's kunnen voorkomen. Het {path} kan dan gebruikt worden om extra namespaces te creëren.

Aanbevelingen voor het {domain}

1. Één taak: het register

Het {domain} is bij voorkeur exclusief gereserveerd voor publicatie van het register en het resolvable van de URLs van het register. Als het domein namelijk een onderdeel is van een uitgebreider domein, waarop ook nog andere publicaties plaatsvinden, dan kan er vroeger of later sprake zijn van de noodzaak tot herorganisatie van de publicaties, met alle gevolgen van dien voor de persistentie van de URLs in het register.

2. Geen organisatienaam in het {domain}

Het is sterk af te raden om in het {domain} een organisatienaam op te nemen, hoe verleidelijk dat ook vanuit marketing oogpunt kan zijn. Opnieuw is persistentie hierbij het belangrijkste argument. Organisaties kunnen immers gesplitst, gefuseerd, of hernoemd worden en zij krijgen dan doorgaans een nieuwe naam en kiezen een nieuw internetdomein. Het hernoemen van de URLs verstoort de persistentie. Het blijven gebruiken van het oude domein – iets waar puur technisch niets op tegen zou zijn – kan echter de indruk wekken dat de data ook verouderd is. Registers zullen over het algemeen blijven bestaan zolang ze een bepaald nut dienen. Als het register daadwerkelijk wordt opgeheven of overgaat in een nieuw register, dan zijn de modellen en referentieobjecten in het oude register doorgaans ook daadwerkelijk uit de tijd.

3. Terughoudend met `{path}`

Probeer het gebruik van `{path}` zo veel mogelijk te vermijden. Hoe korter de URI, hoe handiger in gebruik. Hoe minder informatie in de URI, hoe kleiner de kans dat er later op teruggekomen moet worden.

`{type}`

Het `{type}` geeft aan om wat voor soort URI het gaat. Dit kan zijn:

`'id'`

identificerend van een object (individual/instance) in een register.

`'doc'`

documentatie (metadata) over het object in het register.

`'def'`

definitie van een term in een ontologie.

Aanbevelingen voor `{type}`

1. Gebruik 303 redirect van de 'id'-URI naar de 'doc'-URI.

In 'Cool URIs for the Semantic Web' [17] wordt in sectie 4.2 [18] uitgelegd hoe dit bedoeld is.

2. Gebruik Hash-URIs voor termen uit het model

In een Linked Data applicatie is het onderscheid tussen model en content soms moeilijk te maken. In een relationele database is dat onderscheid doorgaans duidelijker: tabellen en kolommen geven het model aan en de inhoud van de tabellen vormen de content. In Linked Data kun je echter een klasse ook beschouwen als een instance (namelijk van de klasse `rdfs:Class`). Om de gebruiker van een register meer duidelijkheid te verschaffen over welke termen echt tot het model behoren en welke termen gezien kunnen worden als inhoud van het register, verdient het aanbeveling om de URIs van de eersten als hash-URI (#-URI) te definiëren: `http://{domain}/def#{term}`. Dit heeft als bijkomend voordeel dat de URI `http://{domain}/def` alle termen uit het model oplevert.

In 'Cool URIs for the Semantic Web' wordt uitgelegd hoe dit bedoeld is in sectie 4.1. [19]

Als de ontologie erg omvangrijk is, kan ook gekozen worden om geen gebruik te maken van hash-URIs.

`{concept}`

Het `{concept}` geeft de menselijke lezer een indicatie van het soort concept dat de URI identificeert. Het `{concept}` is belangrijk om twee redenen. Ten eerste kan het een uitkomst bieden als objecten binnen de registratie geen unieke identifiers hebben, maar wel uniek zijn per soort object. Bijvoorbeeld gemeente Utrecht en provincie Utrecht. Ten tweede, en dit is belangrijker, levert het een begrijpelijker URI op. Een menselijke lezer kan vermoeden dat <http://bagregister.nl/id/pand/01010101> de URI van een pand uit de BAG is.

Een mogelijk nadeel van het opnemen van `{concept}` in de URI is dat hiermee betekenis in de URI wordt opgenomen, terwijl betekenisloze IDs over het algemeen eenvoudiger persistent te maken zijn.

Aanbevelingen voor `{concept}`

1. `{concept}` betekent niets.

Het is zeer onverstandig om `{concept}` enige betekenis toe te kennen voor de machine. URIs zijn in technische zin opaque. [20] Het is dus niet zo dat het `{concept}` per se de klasse is waartoe een object behoort. Het helpt alleen de menselijke lezer, bijvoorbeeld de beheerder van een semantisch model, om de URIs te herkennen. [21] en [22]

2. Denk ook bij het kiezen van `{concept}` aan persistentie.

Als het in een registratie denkbaar is dat objecttypen (klassen) van naam veranderen, maar dan nog wel dezelfde klasse vertegenwoordigen, is het niet verstandig dit onderdeel in de URI op te nemen. Neem in dat geval een hogere klasse op. Volgens sommigen betekent het veranderen van het type van een instance per definitie dat er niet langer sprake is van dezelfde instance, maar van een andere instance, van het andere type. Voorbeeld: stel dat het Centraal Orgaan opvang Asielzoekers (COA) wordt omgevormd van zelfstandig bestuursorgaan (zbo) naar agentschap. [\[23\]](#) En dat we als URI van het COA zouden kiezen voor: {domein}/id/zbo/coa. Dan wordt dat na de omvorming {domein}/id/agentschap/coa. Zouden we kiezen voor {domein}/id/organisatie/coa dan hoeven we de URI niet aan te passen, maar kunnen we ook geen onderscheid meer maken tussen de COA als ZBO en de COA als agentschap.

{reference}

De {reference} is de identificerende naam of code van het individuele object. Wat betreft {reference} geeft de URI strategie veel vrijheid, aangezien de eisen in verschillende toepassingen sterk uiteen kunnen lopen. Een {reference} kan zijn: een identificerend nummer, een alfanumerieke code, een woord of naam, etc. Elk register heeft wel een manier om de individuele objecten in de verzameling uniek aan te duiden. Deze unieke aanduiding kan worden opgenomen in de {reference}.

Aanbevelingen voor {reference}

1. Namen of nummers?

Er is vaak discussie over het gebruik van 'betekenisloze' identifiers versus 'betekenisvolle' identifiers. Zolang computers geen bewustzijn hebben is elke URI voor de machine een betekenisloze string. Voor mensen kan ook een betekenisloze string betekenis krijgen (020 wordt veel gebruikt door mensen die het label 'Amsterdam' of 'Ajax' niet willen uitspreken, 013 (Tilburgs poppodium), 9292 (OV-informatie), nummer 14 (Johan Cruijff).

Namen of nummers, voor beiden is wat te zeggen. Nummeren heeft als voordeel dat het nauwkeuriger lijkt en er geen homoniemen voor kunnen komen. Maar je verliest herkenbaarheid en hanteerbaarheid voor mensen, zonder steeds de labels bij de hand te hebben.

- In de praktijk zijn de URIs voor de concepten in vrijwel alle semantische standaarden betekenisvol en bevatten zij doorgaans het volledige label (naam) waarmee de term voor de mens wordt aangeduid (meestal als CamelCase [\[24\]](#) geschreven zodat er geen spaties in voorkomen).
- Bij grote aantallen objecten wordt het ondoenlijk om voor elk object een herkenbare unieke naam te bedenken. We gaan dan - vrijwel vanzelf - nummeren.
- Tussen deze twee uitersten zit een grijs gebied. Voor kleine, stabiele sets met objecten (bijvoorbeeld provincies) is het voordelig om de hele naam in de URI op te nemen, bij iets grotere sets, met meer mutaties, komen vaak lange namen voor die de URI onhandelbaar maken. Het kan dan een oplossing zijn om afkortingen in de URI te gebruiken. [\[25\]](#) en [\[26\]](#)

2. Vermijd vreemde tekens in een URI.

Het beste is om zich te beperken tot lowercase letters, cijfers, en eventueel koppeltekens als scheidingstekens.

Open issues

Een aantal issues hebben we wel gesignaleerd, maar hierover zijn nog geen conclusies getrokken of consensus bereikt. De vraag is ook of dat mogelijk en ook noodzakelijk is. Verschil van inzicht zal er altijd kunnen bestaan en ook dan moet Linked Data kunnen functioneren.

Issue1: URNs vs. URIs

Het W3C geeft als best-practice [\[27\]](#):

1. Use URIs as names for things
2. Use http-URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs. so that they can discover more things.

Dit is op zich een zeer krachtig paradigma: http-URIs maken gebruik van de bestaande infrastructuur van het web en elke browser is in staat om een http-URI op te vragen. Door middel van content negotiation – ook weer standaard http functionaliteit – vertaalt de server de aanvraag op basis van de ID naar het gevraagde formaat en levert een document – de 'useful information' – aan de client onder vermelding van de http-returncode '303-redirect'. Deze redirect is nodig omdat de documenten in verschillende formaten immers elk hun eigen URL hebben, die anders is dan de ID die opgevraagd wordt. En dat nu blijkt in de praktijk slecht begrepen en vaak verkeerd te worden geïmplementeerd.

Daarnaast zien we op tal van domeinen al een standaardisatie plaatsvinden in de identificatie van informatie die online beschikbaar is. Dit is nodig – ook als het geen Linked Data is – omdat informatie op het internet op talrijke manieren kan worden gekopieerd en gecombineerd met andere informatie. Een groot verschil met de oude, papieren wereld waar kopieën spaarzaam gemaakt werden en de documenten waarnaar verwezen werd doorgaans in het dossier voorkwamen. Die standaardisatie heeft (vooralsnog?) zelden de vorm van een http-URI, maar is doorgaans een nummer, oftewel een URN (Unified Resource Name). Het bekendste voorbeeld is het ISBN-nummer voor boeken. Een minder bekend, maar recenter en relevanter voorbeeld is de nieuwe standaard ECLI-code van de EU, waarmee rechterlijke uitspraken worden geïdentificeerd. Dus, denkend aan het uitgangspunt om te hergebruiken wat er al voorhanden is, zouden we kunnen overwegen om met de URNs die al gemaakt worden http-URIs te maken. Met een kleine aanpassing blijft de W3C best-practice intact:

1. Use URNs as names for things
2. Use REST-services as resolvers for those URNs, so that people can look up those names.
3. When someone looks up a URN, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URNs. so that they can discover more things.

Nu is helder dat de URN niets oplevert en dat de resolver een beschrijving van het object teruggeeft. Hiervoor is geen 303-redirect nodig. Deze benadering maakt het tevens mogelijk om verschillende resolvers te maken voor dezelfde URN. Elke resolver kan beslissen welke 'useful information' hij verstrekt over een concept. Een voordeel boven de http-URI die altijd naar één locatie leidt.

Issue 2: Herkenbaar internetdomein

In de Britse strategie is men er van uitgegaan dat alle http-URLs van Linked Data van de Engelse overheid ondergebracht worden onder één hoofddomein: 'data.gov.uk'. Inmiddels klinken er echter geluiden om de indeling van het hele 'gov.uk' domein te gaan herzien, waarmee ook de toekomst van data.gov.uk niet langer zeker is. Een belangrijk advies van een van de bedenkers van de Engelse strategie is dan ook om een domein dat vast onderdeel is van persistente URLs in de wet wordt vastgelegd, omdat die persistentie anders niet gewaarborgd is.

Het idee van een hoofddomein is echter wel degelijk aantrekkelijk. Het zorgt voor grote herkenbaarheid van de URLs, iets wat vertrouwen wekt bij de afnemers van de Linked Data. In Nederland zouden we hiervoor het domein data.gov.nl kunnen gebruiken. Bijkomend voordeel van dit domein is dat het nu nog niet in gebruik is en dus nog geen andere associaties heeft.

Aan de andere kant brengt één hoofddomein ook problemen met zich mee. Het domein moet onderverdeeld worden om deze oplossing schaalbaar te houden. Dat kan door sub-domeinen te definiëren. De sector-indeling is hierboven al verworpen. Maar 'No register, no identifier' geeft de oplossing: elk register zou zijn eigen sub-domein kunnen krijgen: {register}.data.gov.nl. Schaalbaarheid is hiermee gewaarborgd, maar er is natuurlijk wel een administratie van alle registers nodig: Een register van registers.

Op zich is het natuurlijk wel handig om zo'n register te hebben, maar het brengt beheerkosten met zich mee en is een potentieel single point of failure.

In de laatste workshop was de conclusie dat het niet realistisch is om een oplossing totaal afhankelijk te maken van een centrale voorziening die weer door een partij moet worden beheerd. De houders van de registers zouden op die manier afhankelijk worden van deze partij om hun URLs te kunnen munten volgens de URI-strategie. Een centraal register van registers mag dus nooit een onmisbaar onderdeel van het stelsel worden. De registerhouders moeten volledig zelfstandig kunnen beschikken over het domein van hun register.

Elk register moet dus op een eigen internetdomein worden ondergebracht: {register}.nl. Het registerdomein wordt gebruikt voor de namespace en voor de authentieke resolver van het register zelf.

Vervolgens is het wel praktisch om een Register van Registers in te richten. Niet als noodzakelijke component, zonder welke het stelsel niet kan functioneren, maar als handige catalogus (/wegwijzer/gazetteer?) voor ontwikkelaars die op zoek zijn naar registers, resolvers en registerhouders.

Het is denkbaar (maar niet noodzakelijk) dat er door het Register van Registers een resolver wordt ingericht (bijv. {register}.data.gov.nl) die redirect naar de resolver van het register (bijv. naar {register}.nl). {register}.data.gov.nl wordt dan een alternatieve resolver voor de authentieke resolver van het register. Maar het is niet verstandig dit als purl-server te zien.

Issue 3: Mate waarin de strategie geformaliseerd dient te worden

Het ultieme doel is een situatie waarin elk object maar 1 URI heeft. Die URI wordt gemunt door de authentieke registratie van dat object. Voor een aantal gegevens ligt dit vast in de wet. De vraag is of het wenselijk en haalbaar is om zo'n formele basis voor authenticiteit op te nemen in de strategie, of dat we, misschien zelfs beter, kunnen vertrouwen op een organisch proces waarbij de meest gebruikte

bronnen de facto registers worden? We zouden formele criteria waaraan een register moet voldoen als volgt kunnen formuleren:

Een register

- is een Standaard of Authentieke Registratie
- heeft een formele basis voor authenticiteit:
 - de Standaard komt voor op de 'pas toe of leg uit'-lijst met open standaarden [\[28\]](#) van het Forum Standaardisatie
 - de Authentieke Registratie wordt als zodanig aangeduid in Nederlandse of EU-regelgeving
- heeft een registerhouder
- heeft een eigen internetdomein (namespace)
- heeft een eigen URI-patroon dat voldoet aan de nationale URI-strategie

Dit geldt ook voor de URI-strategie zelf: Kan de URI-strategie als standaard worden gedefinieerd? En zo ja, is het dan voldoende om deze standaard op de 'pas toe of leg uit'-lijst van het Forum Standaardisatie te plaatsen of zijn er uitgebreidere maatregelen gewenst, misschien zelfs wetgeving?